REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	September 1995	Quarterly Report - 7/1/95 to 9/30/95

4. TITLE AND SUBTITLE
High-Order Modeling Techniques for Continuous Speech Recognition

5. FUNDING NUMBERS
8547-5 - BU Source #
ONR Grant #: N00014-92-J-1778

6. AUTHOR(S)
Mari Ostendorf

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Trustees of Boston University
881 Commonwealth Ave.
Boston, MA 02215

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This research aims to develop new and more accurate stochasitc models for speaker-independent continuous speech recognition by developing acoustic and language models aimed at representing high-order statistical dependencies within and across utterances, including speaker, channel and topic characteristics. These techniques, which have high computational costs because of the large search space associated with higher order models, are made feasible through a multi-pass search strategy that involves rescoring a constrained space given by an HMM decoding. With these overall project goals, the primary research efforts and results over the last quarter have included: 1) an extensive literature survey of research adaptation; 2) development of a trigram word prediction tool for the use in experiments to estimate the entropy of conversational English; 3) further experimental exploration of dependence tree topology design and extension of the modeling framework to handle continuous observation vectors; 4) initiated work on HMM topology design; and 5) furthered efforts on establishing a baseline HTK recognition system for a task of recognizing the Marcophone natural numbers data, on which we currently achieve 76% word accuracy.

14. SUBJECT TERMS	15. NUMBER OF PAGES
	11
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
unclassified	unclassified	unclassified	

ENCL (1)

Boston University

College of Engineering
44 Cummington Street
Boston, Massachusetts 02215
617/353-2811

Electrical, Computer and Systems Engineering

October 17, 1995

Defense Technical Information Center
Building 5, Cameron Station
Alexandria, Virginia 22304-6145

Dear Sir or Madam,

Enclosed is a copy of the quarterly progress report for ONR research grant No. N00014-92-J-1778, "High-Order Modeling Techniques for Continuous Speech Recognition," for the period from July 1995 to September 1995. In addition, I have enclosed a copy of the FY95 annual report which was sent electronically to ONR late July. Please let me know if I can provide any additional information. I would also be happy to hear any feedback you have about the research.

Sincerely,

Mari Ostendorf
Associate Professor
617-353-5430

High-Order Modeling Techniques for Continuous Speech Recognition

Progress Report: 1 July 1995 – 30 September 1995

submitted to
Office of Naval Research
and
Advanced Research Projects Administration

by
Boston University
Boston, Massachusetts 02215

Principal Investigator

Dr. Mari Ostendorf
Associate Professor of ECS Engineering, Boston University
Telephone: (617) 353-5430

Administrative Contact

Maureen Rodgers, Awards Manager
Office of Sponsored Programs
Telephone: (617) 353-4365

Executive Summary

This research aims to develop new and more accurate stochastic models for speaker-independent continuous speech recognition by developing acoustic and language models aimed at representing high-order statistical dependencies within and across utterances, including speaker, channel and topic characteristics. These techniques, which have high computational costs because of the large search space associated with higher order models, are made feasible through a multi-pass search strategy that involves rescoring a constrained space given by an HMM decoding. With these overall project goals, the primary research efforts and results over the last quarter have included:

- an extensive literature survey of research in adaptation;
- development of a trigram word prediction tool for use in experiments to estimate the entropy of conversational English;
- further experimental exploration of dependence tree topology design and extension of the modeling framework to handle continuous observation vectors;
- initiated work on HMM topology design; and
- furthered efforts on establishing a baseline HTK recognition system for a task of recognizing the Macrophone natural numbers data, on which we currently achieve 76% word accuracy.

As usual, substantial software maintenance and development efforts were also required during this period.

Contents

Principal Investigator Name: Mari Ostendorf
PI Institution: Boston University
PI Phone Number: 617-353-5430
PI E-mail Address: mo@raven.bu.edu
Grant or Contract Title: High-Order Modeling Techniques for Continuous Speech Recognition
Grant or Contract Number: ONR-N00014-92-J-1778
Reporting Period: 1 July 1995 – 30 September 1995

1 Productivity Measures

- Refereed papers submitted but not yet published: 0
- Refereed papers published: 1
- Unrefereed reports and articles: 0
- Books or parts thereof submitted but not yet published: 0
- Books or parts thereof published: 0
- Patents filed but not yet granted: 0
- Patents granted (include software copyrights): 0
- Invited presentations: 0
- Contributed presentations: 0
- Honors received: none
- Prizes or awards received: none
- Promotions obtained: none
- Graduate students supported $\geq$ 25% of full time: 3
- Post-docs supported $\geq$ 25% of full time: 0
- Minorities supported: 2 women

Principal Investigator Name: Mari Ostendorf
PI Institution: Boston University
PI Phone Number: 617-353-5430
PI E-mail Address: mo@raven.bu.edu
Grant or Contract Title: High-Order Modeling Techniques for Continuous Speech Recognition
Grant or Contract Number: ONR-N00014-92-J-1778
Reporting Period: 1 July 1995 – 30 September 1995

2 Project Summary

2.1 Introduction and Background

The goal of this work is to develop and explore novel stochastic modeling techniques for acoustic and language modeling in large vocabulary continuous speech recognition, particularly recognition of spontaneous speech. Although significant advances have been made in recognition technology in recent years, spontaneous speech recognition accuracy is still hardly better than 50%. More casual speaking modes introduce additional sources of variability that require improvements at all levels of the recognition process: signal processing, acoustic modeling, lexical representation and language modeling – both in terms of the baseline stochastic models and the techniques for adapting these models. The challenge of spontaneous speech recognition must be met for applications requiring transcription of meetings, voice mail or archived data, for example, but also because spoken inputs in human-computer communication will become more spontaneous as interfaces become more natural.

In addressing these challenges, the general theme of the research in this project is high-level correlation modeling, i.e. representing correlation of observations beyond the level of the frame or the word to dependencies within and across utterances associated with speaker, channel, topic and/or speaking style. Continuing the ARPA-ONR funded work at Boston University (BU) on segment-based acoustic modeling for speech recognition, the current project builds on the stochastic segment model, algorithms developed for topic-dependent language modeling, and the BU recognition system in general. The recognition framework also includes a multi-pass search strategy to accommodate the higher-order (and therefore more computational) models explored here. In particular, we will concentrate on three problems: development of hierarchical models of intra-utterance correlation of phones and model states, e.g. by extending the theory of Markov dependence trees; unsupervised adaptation of acoustic models within and across utterances based on these models; and sub-language modeling triggered by acoustic and dialog-level cues.

The research approach is to develop formal models of statistical dependence that overcome limitations of existing models, in combination with exploring fast search and robust parameter estimation techniques to address the added complexity of these models. By considering radically new but formal models, rather than minor variations of existing models or heuristic patches, this

work offers the potential to address many of the most difficult problems in speech recognition, including recognition of spontaneous speech. By also building on the existing strengths of speech recognition technology, both in the theoretical foundation and in the use of multi-pass search, this work has the added advantage that advances will be more apparent and more easily transitioned to existing systems.

Over the past year, the focus of this project was in three main areas. First, standard n-gram training and dynamic cache language modeling techniques were extended for use in sentence-level mixture modeling [1], yielding a significant reduction in perplexity though only small gains in recognition performance as yet. Second, an algorithm was developed and implemented for training discrete dependence trees with missing observations [2]. Initial experiments explored tree topology design issues and obtained improved prediction error using dependence trees in a simple adaptation experiment. Third, lattice search algorithms were implemented to reduce computation in segment model rescoring, including a local search algorithm suitable for the higher-order language and acoustic models explored in this work [3]. Other results included development of a parametric segment model clustering algorithm and exploration of auditory-based signal processing algorithms. The research efforts were coordinated with another project involving channel modeling for improved telephone speech recognition. In addition to these research advances, significant effort was devoted to software system improvements and participation in the ARPA speech recognition benchmarks, where BU achieved 11.6% word error in the officially reported result and 10% word error using the BBN benchmark system for first pass scoring [4].

In the previous quarter, we have continued to build on these results, and also started some new thrusts, as described in the next section.

2.2 Summary of Recent Work

The research efforts during this period, supported in part by AASERT awards from ONR and ARPA, covered a variety of problems, as summarized below.

Adaptation Literature Review As part of our efforts to develop a new approach to incremental adaptation, we conducted an extensive literature survey of adaptation techniques for HMMs. As a result of this survey, we have a good understanding of the underlying assumptions and limitations of current approaches, which will help in development of our own alternative approach. This review will be included in Ashvin Kannan's thesis proposal, which is in preparation.

Measuring the Entropy of Conversational Speech. In order to better understand the potential for using acoustic cues in language modeling, we are conducting a text prediction experiment with human subjects using Switchboard conversations. Because an earlier version of the experi-

ment showed that humans did not do as well as the trigram at predicting words, Rukmini Iyer is integrating a trigram prediction tool into the prediction game interface to aid the human subjects and hopefully yield a better estimate of the entropy of conversational speech.

Intra-utterance phoneme dependence modeling. We have extended our previous work in dependence tree models by drawing analogies to different HMM mixture distribution assumptions. This led to the use of phone-specific codebooks, which gave more robust estimation and better prediction results on independent test data. In experiments with phone vs. speaker state vectors, Orith Ronen developed additional insight into robust tree topology training, and she is further exploring this issue in experiments on the WSJ corpus. (Earlier experiments were on the TIMIT corpus.) In addition, we have extended the theoretical modeling framework to use the tree as a hidden state vector that models continuous cepstral vectors.

HMM Topology Design. Currently, most recognition systems use a fixed number of HMM states for all acoustic sub-word models (e.g. triphones), ignoring the fact that some phones are much shorter than others. Our goal is to design context-dependent HMM topologies, building on a maximum-likelihood variation of successive state splitting [5], that can represent reduction phenomena that are so problematic in spontaneous speech. As a first step in this effort, Song Xing is implementing software and running Switchboard benchmark experiments that make use of HTK tools.

Recognition of telephone speech. In an ARPA-sponsored project coordinated with this effort, we have been developing a baseline system for recognizing word strings with natural numbers, based on a subset of the Macrophone corpus [6]. In the past quarter, Rebecca Bates has focused on improving the baseline recognition performance. Through changes to the dictionary and grammar, word error rates were reduced from 28% to 24%. We are in the process of retraining the model on the entire Macrophone corpus, as well as moving to a bigram grammar, and we expect this to lead to a significant reduction in error. The baseline system will be evaluating different channel modeling algorithms.

Software Maintenance As in most quarters, some effort was devoted to software maintenance. In particular, all software was updated to handle a compiler upgrade, and other software has been rewritten to increase efficiency. In addition, substantial effort was devoted to transferring the recognition decoder software expertise to a new student, Ashvin Kannan, since the author of that software, Fred Richardson, graduated this past year.

2.3 Future Goals

The goals of this project in the next quarter are:

- investigate segment-level trajectory clustering and improve distribution clustering through more general context conditioning;
- implement the extension of the dependence tree model that represents continuous observations using the discrete tree as a hidden state;
- re-assess dynamic language model improvements using a cleaner version of the NAB training data; and
- evaluate the current system and anticipated advances on the Switchboard spontaneous speech recognition task.

References

[1] R. Iyer, M. Ostendorf and J. R. Rohlicek, "Language Modeling with Sentence-Level Mixtures," *Proc. ARPA Workshop on Human Language Technology,* March 1994, pp. 82-87.

[2] O. Ronen, J. R. Rohlicek and M. Ostendorf, "Parameter Estimation of Dependence Tree Models Using the EM Algorithm," *IEEE Signal Processing Letters,* Vol. 2, No. 8, August 1995, pp. 157-159.

[3] F. Richardson, M. Ostendorf and J. R. Rohlicek, "Lattice-based Search Strategies for Large Vocabulary Speech Recognition," *Proc. Int'l. Conf. on Acoust., Speech and Signal Proc.,* pp. 576-579, 1995.

[4] M. Ostendorf, F. Richardson, R. Iyer, A. Kannan, O. Ronen and R. Bates, "The 1994 BU NAB News Benchmark System," *Proceedings of the ARPA Workshop on Spoken Language Technology,* January 1995, pp. 139-142.

[5] M. Ostendorf and H. Singer, "A Maximum Likelihood Variation of Successive State Splitting for HMM Topology Design," manuscript in preparation.

[6] K. Taussig and J. Bernstein, "Macrophone: An American English Telephone Speech Corpus," *Proc. ARPA Spoken Language Technology Workshop,* 1994.

Principal Investigator Name: Mari Ostendorf
PI Institution: Boston University
PI Phone Number: 617-353-5430
PI E-mail Address: mo@raven.bu.edu
Grant or Contract Title: High-Order Modeling Techniques for Continuous Speech Recognition
Grant or Contract Number: ONR-N00014-92-J-1778
Reporting Period: 1 April 1995 – 30 June 1995

3 Technical Transitions

The initial grant included a subcontract to BBN, and BU has collaborated with BBN by combining the Byblos system with the SSM in N-Best hypothesis rescoring to obtain improved recognition performance, providing BBN with software as well as papers and technical reports to facilitate sharing of algorithmic improvements. In addition, BU student Rukmini Iyer worked at BBN as part of a graduate student co-op program, and she also participated in the 1995 Workshop on language modeling at Johns Hopkins University.

The recognition system that has been developed under the support of this and other grants was also used for obtaining phonetic alignments for a corpus of radio news speech collected at BU with partial support from the Linguistic Data Consortium.

More generally, the results of this work are of interest to the speech research community and have been made available through timely dissemination in papers and presentations. The students trained on this grant also serve to transfer technology when they graduate.

Principal Investigator Name: Mari Ostendorf
PI Institution: Boston University
PI Phone Number: 617-353-5430
PI E-mail Address: mo@raven.bu.edu
Grant or Contract Title: High-Order Modeling Techniques for Continuous Speech Recognition
Grant or Contract Number: ONR-N00014-92-J-1778
Reporting Period: 1 July 1995 – 30 September 1995

4 Publications and Presentations

During this reporting period, we published one journal paper.

Refereed papers published:

"Parameter Estimation of Dependence Tree Models Using the EM Algorithm," O. Ronen, J. R. Rohlicek and M. Ostendorf, *IEEE Signal Processing Letters,* Vol. 2, No. 8, August 1995, pp. 157-159.

On-line information:

General information about the research in the Signal Processing and Interpretation Laboratory (SPI Lab), headed by Prof. Ostendorf, is available by browsing the SPI Lab WWW home page (http://raven.bu.edu/), which includes a description of this and related projects and a publication list. Technical reports and recent theses can be obtained by anonymous ftp to raven.bu.edu (in the pub/reports directory).

Principal Investigator Name: Mari Ostendorf
PI Institution: Boston University
PI Phone Number: 617-353-5430
PI E-mail Address: mo@raven.bu.edu
Grant or Contract Title: High-Order Modeling Techniques for Continuous Speech Recognition
Grant or Contract Number: ONR-N00014-92-J-1778
Reporting Period: 1 June 1995 – 30 September 1995

5 Team Members

- Principal Investigator: Mari Ostendorf
- Graduate students:
  - Orith Ronen, Ph.D. candidate
  - Ashvin Kannan, Ph.D. candidate
  - Rukmini Iyer, M.S. 1994, Ph.D. candidate
- Undergraduate students
  - Greg Grozdits, B.S. candidate
- Visiting researcher: Song Xing (partial support)

This project is coordinated with work funded by an ARPA AASERT award on channel modeling for speech recognition which supported graduate student Rebecca Bates.